Chapter 8

MUSICAM And Psychoacoustic Parameters

The patented Psychoacoustic Parameter adjustment capabilities enables the user to tailor the enhanced Layer 2 MUSICAM algorithm to achieve the best possible audio under all conditions.

8. Psychoacoustic Parameter Adjustments for MUSICAM

8.1 Introduction

With a bandwidth of 20 Hz-20,000 Hz, and a signal-to-noise ratio better than 92 dB, the digital capacity requirement for a high quality stereo audio signal is approximately 1.4 Mbps (million bits per second) (2 x 706 kb/s for a 44.1 kHz, 16 bit PCM signal). This format is frequently used by and recommended for professional and consumer equipment by a number of international bodies such as CCIR and IEC. One of the most used product types with this format is the Compact Disc. Due to the direct relationship between bit rate and costs for transmission and storage it is desirable to reduce the rate.

In recent years, many different low bit rate audio coding schemes have received much attention, such as ADPCM methods. The bit rate reduction achieved by these coding processes is mainly due to the reduction of redundancy using statistical correlation. In addition, the more sophisticated coding schemes exploit psychoacoustical phenomena, such as the ears masking functions in time and frequency domain to achieve lower bit rates.

Every audio signal contains irrelevant signal components which are not responsible for the identification of the audio signal; i.e., for the determination of timbre and localization. The information process in the brain does not require these irrelevant signals. The reduction of irrelevance means that these signal components are not transmitted, which results in a low bit rate without any perceived degradation of the audio quality. Furthermore, it is possible to allow a certain degree of quantizing noise, which is inaudible to the human ear due to the masking effects of the audio itself.

To understand this masking effect, the concept a masking tone must be defined. A masking tone, often called a “masker”, is simply a high amplitude audio signal occurring over a relatively narrow frequency span. Typically, an audio signal contains a number of these masking tones occurring at several different frequencies.

A masking tone renders smaller amplitude tones close to it inaudible due to its masking effect. The exact shape of the masking effect is called the masking threshold. The aggregate of all the maskers defines a global masking threshold and the parts of an audio signal below the global masking threshold are inaudible. They are said to be masked and therefore need not be transmitted. Other signal components above the masking threshold require only the level of quantization to keep quantization noise below the masking threshold, and thus the quantization-induced noise remains inaudible.

Quantization noise can be better adapted to the masking threshold of the human ear by splitting the frequency spectrum into sub-bands. The quantization of the analog time samples required for each sub-band is dependent on the minimum masking value in each sub-band. This minimum masking level is a measure of the allowed quantization noise that is just below the level of perceptibility. Sub-bands whose desired signals are well below the masking threshold (and are thus irrelevant for the human ear) do not need to be transmitted.

In each 24 millisecond period (when using 48k sampling rate), a calculation of the masking threshold is performed for each sub-band. This threshold is then used to compute the psychoacoustically best allocation of the available bits. This process is called dynamic bit allocation. Audio data is quantized using the dynamic bit allocation and thus the required bit rate for time-variant audio signals changes continuously due to the changing masking threshold. If there is an insufficient number of bits to hide the quantizing induced noise completely, then the noise is placed in the least objectionable place in the audio sample. If there is an excess number of bits, then the extra bits are used to reduce the quantizing induced noise to the lowest possible level. The allocation of the extra bits is crucial and allows multiple encode-decode cycles as well as post production of the audio. **It is therefore best to use the highest transmission bit rate available for critical applications.** In addition, your codecs inherently wide dynamic range provides extra noise immunity.

The total transmitted bit stream contains quantized audio values as well as auxiliary information describing bit allocation and scale factors, all of which are required by the decoder to reproduce the audio information.

The scale factors are determined by searching for the maximum sampling value in each sub-band and quantizing the result using 6-bit sampling. The scale factors have a dynamic range of 120 dB and are thus sufficient for future encoding for quantized PCM signals using up to 20-bit sampling while still retaining their dynamic range. All necessary information is encoded into MUSICAM frames, each of which represents about 24 milliseconds of real-time audio.

8.2 History

The recent decade has brought improvements in the area of high quality audio transmission and storage. More and more signal processing power at the developers disposal has resulted in real-time perceptual coding techniques. The MPEG I and MPEG II audio standards follow a simulation of the human sound perception for the encoding process. The bit rate needed for transmission or storage of high quality audio signals (1,412 kb/s for Compact Disc) has been reduced to about 200 kb/s as a result of major progress in the development of source coding techniques that utilize knowledge of the auditory system. This means that the average quantization of the audio signal at a sampling frequency of 44.1 kHz would be approximately 2 bits per sample instead of 16 bits per sample as used in CD’s. Despite this high reduction in bit rate, no quality differences are detectable, even by well trained listeners.

With MPEG encoding, real-time audio transmission is now possible via available multiples of 64 kb/s and 56 kb/s digital transmission channels. The relatively low cost of ISDN and satellite channels now lessens the need for high bit rate channels and allows economical use of 192 kb/s and higher bit rates. Utilizing higher channel rates results in higher quality audio.

The knowledge of how the human ear perceives sound is not new in the field of audio coding. The 3 kHz bandwidth of the telephone was chosen because of the source and sink, i.e., the human auditory and speech system. It has been known for a number of years that the main format areas of speech, which result in intelligibility and speaker recognition, are in the frequency range of 300 Hz to 3 kHz. To store or transmit audio signals efficiently, the source and sink must match the transmission system in order to transmit and/or store audio economically.

Before the MPEG (Motion Pictures Experts Group) began the design of a high quality coding algorithm for video and audio, algorithms such as CCITT G.722 were well known and used frequently for audio transmission. The G.722 algorithm provides a bandwidth of <7.5 kHz and a signal-to-noise ratio of 78 dB at 48, 56 and 64 kb/s transmission rates.

Today's audio coding implementations apply psychoacoustic masking models which are based on results of investigations which were performed more than 30 years ago. Recent research in psychoacoustics have only updated and refined the older theories. The most important advance in wide bandwidth coding has been achieved with MPEG 1. This standard provides algorithms for high quality audio coding at bit rates from 32 to 448 kb/s.

8.3 Basic Principle of MUSICAM

The basic principle of MUSICAM is to divide the spectrum of the digitally sampled wide band signal into a definite number of subbands using a suitable filter bank. Adapted to the ears masking effects in time, successive samples of each subband are combined to a block. A block compounding technique is applied to each subband. Once in each block, the peak value is represented by a scale factor (SCF) which has to be quantized. The global masking threshold of the audio signal is estimated using a parallel calculated FFT and the quantization of each subband is determined by the actual calculated signal-to-mask ratio in each subband for a certain time interval. This is the so-called dynamic bit allocation (BAL). The samples of three successive blocks in each subband (3 blocks = 1 frame) are quantized using this dynamic bit allocation which is derived by the ratio of the maximum signal level and the minimum of the masking threshold within each subband.

Every audio signal produces a masking threshold in the ear depending on a time and frequency varying function of the signal. Masking has been defined as “the process by which the threshold of audibility for one sound is raised by the presence of another sound” (American Standards Association, 1960).

The results of masking threshold measurements for narrowband signals (considering tone masking noise and, vice versa, the dependency on the frequency and the level) are the basis for an accurate estimation of the masking threshold of a time variant complex audio signal.

Fig. 8-1 shows the masking threshold generated by a 1 kHz sine wave masker. Signal components below the masking threshold are masked and therefore not audible.

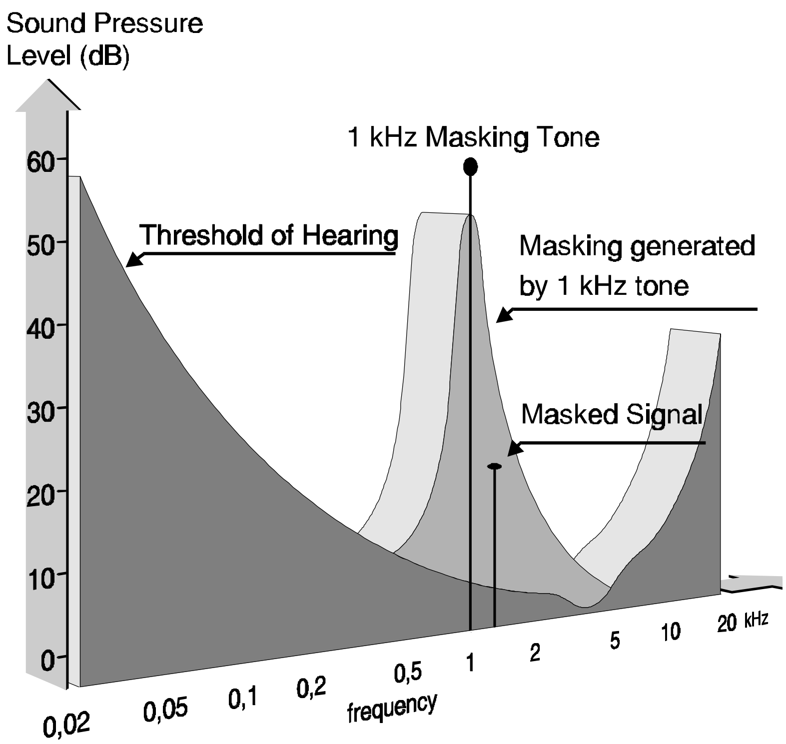


Figure 8-1 Masking threshold generated by a 1 kHz sine wave masker

8.3.1 MPEG Encoder

The block diagram of the MPEG Layer II encoder is shown in Fig. 8-2. The coding principle of MPEG Layer II is subband coding. A polyphase filter network is used for dividing the broadband audio signal into 32 subband signals with a constant bandwidth. With a sampling frequency f_S of the input audio signal, each subband is sampled at $f_S/32$ and has a bandwidth of $f_S/64$ (e.g., at f_S = 48 kHz, $f_S/32$ = 750 Hz). Considering the necessity for both a perfect reconstruction of the signal and achieving a total cancellation of the aliasing distortions in the synthesis filter bank, an optimized filter with a polyphase matrix and a low complexity with a structure leading to a low delay seems to be the best way.

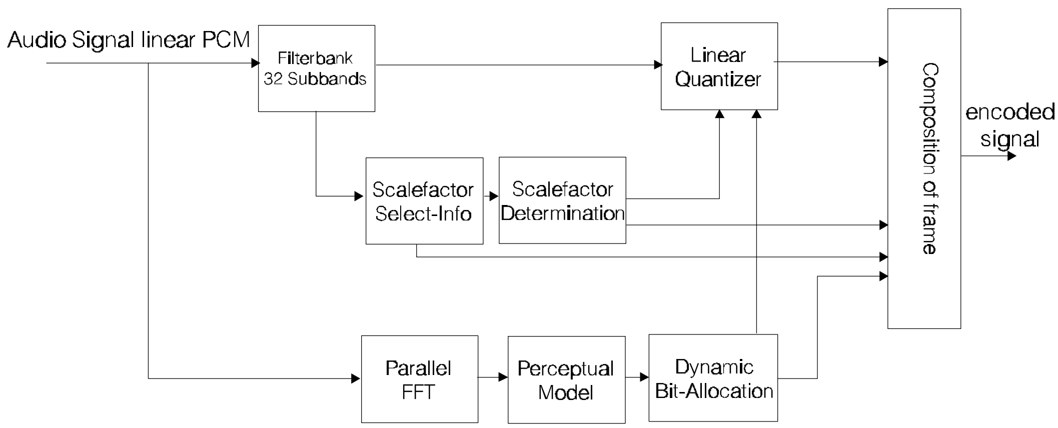


Figure 8-2 Block diagram of the ISO/MPEG Layer II encoder

In each subband, 12 subband samples are combined to a block with a duration of 8 ms. at a sampling frequency of 48 kHz. This was found to be an optimized length considering the worst case of the temporal masking effects, i.e., the pre-masking of quantization noise arising just before the audio signal. In each block of each subband, the peak value of the 12 successive samples is determined and quantized as a 6 bit scale factor with a dynamic range of about 120 dB. The scale factors guarantee an optimal exploitation of the actual dynamic range of the quantizer.

The bit rate of the scale factors is reduced by the elimination of irrelevance and redundancy using a special coding technique. Normally, a constant decay of the scale factors towards higher subbands can be found, similar to the typical spectral slope of speech and music signals. In addition, there is only a small deviation of successive scale factors in time. The probability of a difference of successive scale factors which is more than 2 dB is less than 10%. To exploit this behavior the three successive scale factors of each subband within one frame are considered and classified into defined patterns. Depending on the pattern, one, two or all three scale factors are transmitted together with a so-called scale factor select information, which indicates the pattern. If there are only small deviations between the following scale factors, only one scale factor has to be transmitted. Considering an attack, e.g., triangle, two or even all three scale factors have to be transmitted. This technique reduces the scale factor bit rate from 15 kb/s to about 7 kb/s per channel. Knowledge about the small deviations between successive scale factors may also be used for concealment techniques.

Determination of the bit allocation requires an analysis of the audio signal that is as accurate as possible. Therefore, the uncertainty of the spectrum estimation performed by the filter bank (750 Hz constant bandwidth at f_S = 48 kHz) is compensated by calculating, in parallel addition, a Fast Fourier Transform. This analysis is necessary only in the encoder, and it should be noted that only the samples of the subbands are coded and transmitted. The 1024 point FFT is processed every frame and calculates 512 spectral samples with a spectral resolution of 1/8 of the subbands bandwidth (46 Hz at f_S = 48 kHz). Using such a transform, which allows a high frequency resolution, the human ears masking thresholds can be estimated in time, using the scale factors and subband samples, and in frequency, using the spectral samples, with a very high accuracy.

In each frame period, the calculated masking threshold is used to compute the psychoacoustically best allocation of the available number of bits, depending on the available bit rate of the transmission channel or storage media. For each of the 32 subbands, the minimum of the masking threshold, which determines the maximum level of the just allowed imperceptible quantizing noise is calculated. The necessary resolution of quantizing the subband samples in an individual subband can be derived directly from the difference between the maximum level and the minimum of the masking threshold within each subband. No transmission capacity is necessary for those subbands which are completely masked by much more important components of adjacent subbands. Experiments have shown that the bit allocation of each subband signal may vary significantly from one frame to another frame only in the case of attacks in the audio signal, e.g., castanets or drums.

The total transmitted bit stream contains quantized audio samples as well as control information describing bit allocation, scale factor select information and scale factors, all of which are required by the decoder to reproduce the audio information.

The constantly changing dynamic bit allocation leads, first of all, to a *dynamic* bit rate of the encoded audio signal. Most applications require a constant bit rate; thus, the dynamically-varying bit rate margin can be used for different purposes:

- The mask-to-noise ratio can be increased by allocating more bits than necessary. This allows multiple encode-decode cycles as well as post processing of the audio. This is the typical procedure using transmission and storage channels with a fixed capacity.

- A dynamic error protection can be applied, such that during silence when the signal has a low bit rate and is very sensitive to bit errors, a high amount of redundancy can be used for the protection. During the time a high bit rate is required, only a low protection can be given to the audio with the advantage that errors could be masked by the audio signal itself.

- Non-time-critical additional data, like traffic or program information, can be transmitted.

The MPEG Layer II compression algorithm has been designed to take advantage of future advances in psychoacoustic research. The bitstream is defined and thus also the decoder. The *encoder* performs *all* of the psychoacoustic calculations of the algorithm. This technique allows the system to be upgraded by simply changing the encoder software. The encoder enhancements will then be reflected at the output of *all* decoders.

8.3.2 MPEG Decoder

The block diagram of the Layer II decoder is shown in Fig. 8-3. First, the encoded subband samples and separates the control information, i.e., bit allocation, scale factor select information and scale factors are separated from the received multiplex signal. The reconstruction process is characterized by the expansion of the encoded subband samples using the scale factors and the bit allocation for each subband. Using a filter network, which is completely inverse to the filter used in the encoder, the audio signal is reconstructed.

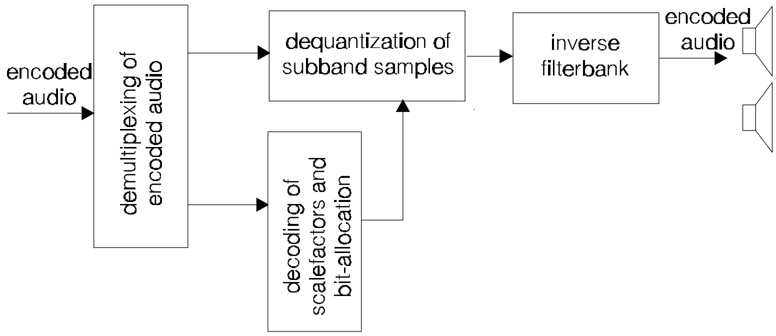
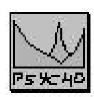


Figure 8-3 Block diagram of the ISO/MPEG Layer II Decoder

The decoding process requires significantly less computational power than the encoding process. The relation is about 1:3. Due to the low complexity and the straightforward structure of the algorithm, it can be easily implemented into a special VLSI chip and the availability of such dedicated VLSIs is high.

8.4 Adjustment of the Perceptual Model—A Tool for the Sophisticated User

All models of ***CDQPrima*** and the ***Prima LT Plus*** allow adjustment of the psychoacoustic parameters from the front-panel keypad. Psychoacoustic parameters can only be adjusted on the ***Prima LT*** from an attached terminal or from the Windows Remote Control program.

<Encoder><Psycho><Tbl Num> <System Setup><Adv Encoder><Psycho><Tbl Num>	EPD		Get default Psychoacoustic parameter table number
<Encoder><Psycho><Load Tbl> <System Setup><Adv Encoder><Psycho><Load Tbl>	EPL		Load psychoacoustic parameters from flash
<Encoder><Psycho><Reset> <System Setup><Adv Encoder><Psycho><Reset>	EPB		Load default psychoacoustic parameters
<Encoder><Psycho><Set parm> <System Setup><Adv Encoder><Psycho><Set parm>	EPP		Set psychoacoustic parameter
<Encoder><Psycho><Store Tbl>	EPS		Store psychoacoustic parameters in flash

<System Setup><Adv Encoder><Psycho><Store Tbl>			
<Encoder><Psycho><Assign Tbl> <System Setup><Adv Encoder><Psycho><Assign Tbl>	EPT		Assign psychoacoustic parameter table
N/A	EPY		Set psychoacoustic parameter type
<Common><General><Set dflts> <System Setup><Defaults>	CDF	Setup, Default	Set default parameters

The "tuning" of a codec has long been a secret of codec manufacturers. Now, with the ***CDQPrima*** and ***Prima LT*** platforms, the user may adjust these tuning parameters. The benefit of this feature is that it allows users to control their own audio quality. This independence from codec manufactures is vital to achieve higher quality levels. The more ears that listen to an algorithm and tune it, the higher the audio quality.

There are 32 psychoacoustic parameters that control the codec (the original ISO MPEG Layer II standard specifies only 10.) These parameters are numbered 0 through 31 and are set by the **EPP** command. Each of the 32 parameters can be one of four types. These are dB, Bark, floating point and integer. The type of each parameter is set by the **EPY** command and should not be changed. The **EPY** command values are set by the **CDF** command.

Your codec offers 20 different compressed digital audio bit rates and 6 sampling rates. Since each bit rate / sample rate combination requires different parameter values, this makes a total of 120 different psychoacoustic parameter tables. There are 240 available tables, which are numbered 0 to 239. The tables from 0 to 119 hold user defined parameters while the tables from 120 to 239 hold the factory defined tables. The tables from 120 to 239 should never be changed but can be copied into the working area (by the **EPL** command), modified (by the **EPP** command) and saved and modified in a user defined table (by the **EPS** command).

TBL	BR
120	8
121	16
122	24
123	32
124	40
125	48
126	56
127	64
128	80
129	96
130	112
131	128
132	144
133	160
134	192
135	224
136	256
137	320
138	384
139	399

Table 8-1
16 kHz

TBL	BR
140	8
141	16
142	24
143	32
144	40
145	48
146	56
147	64
148	80
149	96
150	112
151	128
152	144
153	160
154	192
155	224
156	256
157	320
158	384
159	399

Table 8-5
22.05 kHz

TBL	BR
160	8
161	16
162	24
163	32
164	40
165	48
166	56
167	64
168	80
169	96
170	112
171	128
172	144
173	160
174	192
175	224
176	256
177	320
178	384
179	399

Table 8-4
24 kHz

TBL	BR
180	8
181	16
182	24
183	32
184	40
185	48
186	56
187	64
188	80
189	96
190	112
191	128
192	144
193	160
194	192
195	224
196	256
197	320
198	384
199	399

Table 8-3
32 kHz

TBL	BR
200	8
201	16
202	24
203	32
204	40
205	48
206	56
207	64
208	80
209	96
210	112
211	128
212	144
213	160
214	192
215	224
216	256
217	320
218	384
219	399

Table 8-2
44.1 kHz

TBL	BR
220	8
221	16
222	24
223	32
224	40
225	48
226	56
227	64
228	80
229	96
230	112
231	128
232	144
233	160
234	192
235	224
236	256
237	320
238	384
239	399

Table 8-1
48 kHz

When you set the encoder to operate at a specified sampling rate and bit rate, the codec automatically loads the corresponding psychoacoustic table into the encoder. The current table number used for each sampling rate and bit rate can be displayed or changed by the **EPT** command.

To modify a factory default table, store it in a user table and tell the encoder to use the new table, proceed as follows:

1. Find the default table number for the desired sampling and bit rate by the **EPD** command. The number returned ranges between 120

and 239 and is the psychoacoustic table number used for the specified sampling and bit rate (called the *default table number*). Two numbers are returned by this command, the second number is usually used as the table number to store the modified table into (called the *suggested new table number*).

2. Execute the **EPL** command with the *default* table number to read the psychoacoustic table into memory and to download it to the encoder DSP.

3. Modify the individual psychoacoustic parameters with the **EPP** command until the desired audio quality is achieved.

4. Store the modified table in the *suggested* new table number by the **EPS** command.

5. Tell the encoder to use this new table for the specified sampling and bit rate by executing the **EPT** command.

It is possible and often useful to use the **EPT** command to assign multiple sampling and bit rates to the same table to minimize the table building error.

8.5 Changing Psychoacoustic Parameters

MUSICAM enhanced MPEG Layer II is based on the Psychoacoustic Model I as described in the Annex of the ISO11172-3 document. This model assumes that there are two kinds of maskers, tonal and noises. The tonal maskers arise from signals that generate nearly pure tones or signals that are harmonically rich. Some instruments may generate many harmonics, but each harmonic is relatively pure (narrow in width). Fig. 8-4 shows a tonal masker. The tonal masker is so pure that a single vertical line in this figure represents it. The frequency resolution in the MPEG Layer II psychoacoustic model is 48,000/1024 Hz wide, or about 46 Hz, so the line in Fig. 8-4 represents 46 Hz of bandwidth. The peak power of the tonal masker is called *P*.

An instrument that produces many harmonics, such as a violin or a trumpet, may have many such tonal maskers. The definition of how to identify a tonal masker is described in the ISO specification.

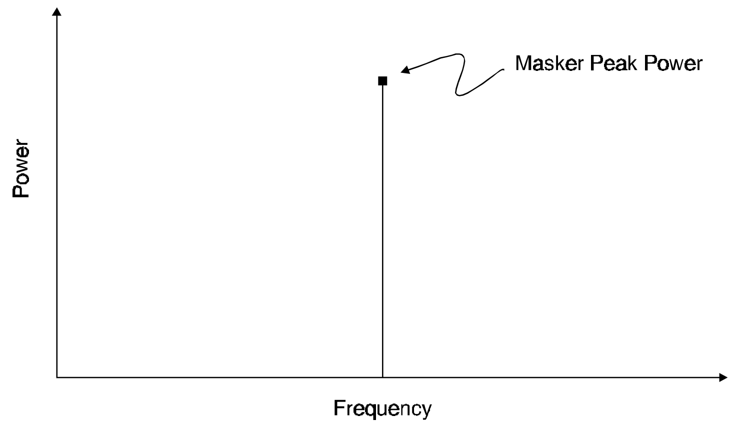


Figure 8-4 Tonal masker

Fig. 8-5 shows a tonal masker with its associated masking skirts. A signal which falls below the masking skirt (such as A) cannot be heard because the of the masking of the larger tone while a smaller amplitude tone (such as B) can be heard if it is above the masking skirt.

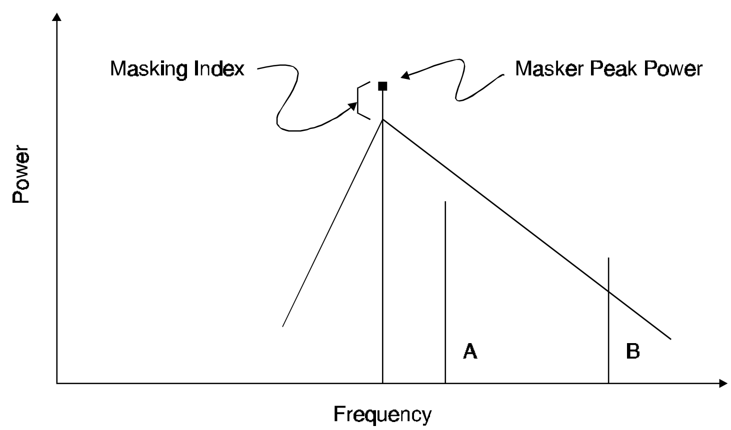


Figure 8-5 Tonal masker with masking skirts

The masking index shown in Fig. 8-5 is the distance from the peak of the tonal masker to the top of the masking skirt. This distance is in dB, and is frequency dependent. The frequency in psychoacoustic measurements is often measured in Bark instead of Hertz. There is a simple function which relates Bark to Hertz. The frequency scale in Hertz which goes from 0 to 20,000 is approximately 0 to 24 Bark. The Bark - Hertz mapping is highly nonlinear.

Fig. 8-6 shows the actual masking skirt as described in the ISO specification for Psychoacoustic Model I. Notice that various slopes of the masking skirt depend on the level of the masker. The masking index, AV, is a function of frequency. These are well known characteristics, which have been determined by psychoacoustic studies.

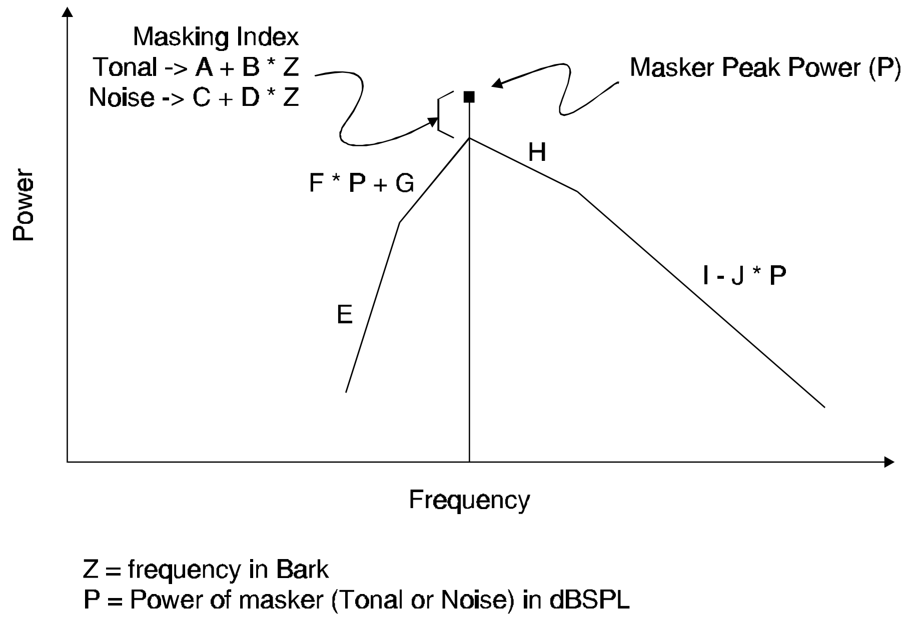


Figure 8-6 ISO Psychoacoustic Model I definitions

Where:

Z = frequency in Bark
DZ = distance in Bark from masker peak (may be + or -)
Pxx(Z(k)) = Power in SPL (96 dB = +/-32767) at frequency Z of masker k
    xx = tm for tonal masker or nm for noise masker
    Pxx is adjusted so that a full scale sine wave (+/- 32767) generates a Pxx of 96 dB.
Pxx = XFFT + 96.0 where XFFT = 0 dB at +/-32767 amplitude
    XFFT is the raw output of an FFT. It must be scaled to convert it to Pxx
AVtm(k) = A + B * Z(k) Masking index for tonal masker k
AVnm(k) = C + D * Z(k) Masking index for noise masker k
VF(k,DZ) = E * (|DZ| - 1) + (F * X(Z(k)) + G) -3 <= DZ < -1
VF(k,DZ) = (F * X(Z(k))+ G) * |DZ| -1 <= DZ < 0
VF(k,DZ) = H * DZ 0 <= DZ < 1
VF(k,DZ) = (DZ - 1) * (I - J * X(Z(k))) + H 1 <= DZ < 8
MLxx(k,DZ) = Pxx(k) - (AVxx(k) + VF(k,DZ))
MLxx is the masking level generated by each masker k at a distance DZ from the masker.
Pxx = Power for tm or nm
    xx = tm or nm

Sometimes audio has no single dominant frequencies (tonals) but is more noise like. In this case, a noise masker is constructed by summing all the energy within 1 Bark (a critical band) and forming a single "noise" masker at the center of the critical band. Since there are 24 Bark (critical bands) then there are 24 noise maskers. The noise maskers are treated just like the tonal maskers. This means that they have a masking index and a masking skirt. Audio may or may not have tonal maskers but it will *always* have 24 noise maskers.

The basic idea is to adjust the model parameters (A, B, ...) until all audio material sounds good. This process of tuning the model parameters is best done when the demand bit rate is used. This demand bit rate is the exact bit rate defined by the model and may usually vary with time. When tuning using the demand bit rate mode, it is important the demand bit rate is observed. The model parameters should be adjusted for the best sound with the minimum demand bit rate. Once the parameters have been optimized in the demand bit rate mode, they can be confirmed by running in the constant bit rate mode.

Although you can use the keypad or a remote terminal to adjust any parameter, we recommend that the Windows Remote Control programs be used. This program enables you to adjust all parameters using sliders and radio buttons, and also gives graphical indications of the relative values of each parameter. A sample screen is shown here:

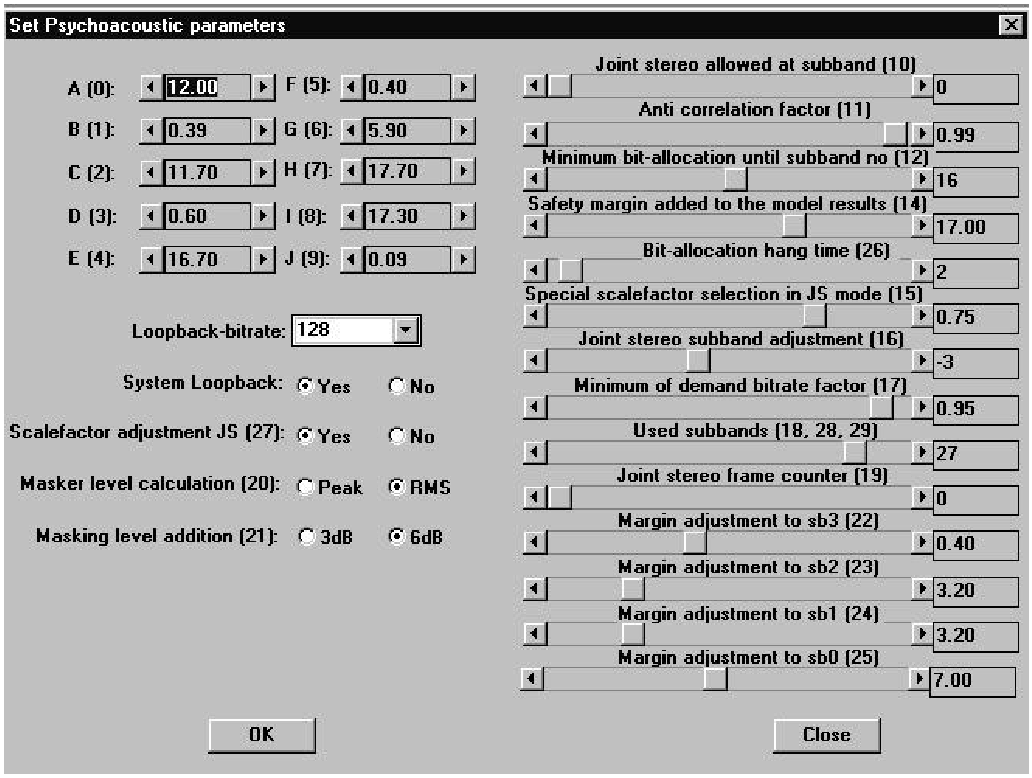


Figure 8-7 Psychoacoustic parameter adjustment screen

8.6 MUSICAM USA psychoacoustic model parameters

MUSICAM USA utilizes the same psychoacoustic model as described in the ISO Psychoacoustic Model I with several additional parameters to enhance the audio quality.

Psychoacoustic parameters A through J are shown in Fig. 8-6 and are described below.

8.6.1 Parameter A - J (0 - 9)

The number enclosed between the parenthesis is the psychoacoustic number referred to by the **pp** argument of the **EPP** command.

	ISO Spec	CCIR (12/92)
A (0) =	6.025 dB	9.0
B (1) =	0.275 dB/Bark	0.3
C (2) =	2.025 dB	5.0
D (3) =	0.175 dB/Bark	0.3
E (4) =	17.0 dB/Bark	17.0

F (5) =	0.4	1/Bark	0.4
G (6) =	6.0	dB/Bark	6.0
H (7) =	17.0	dB/Bark	20.0
I (8) =	17.0	dB/Bark	17.0
J (9) =	0.15	1/Bark	0.1

The above parameter values have been defined in the specifications. Parameter adjustments K through | (10 - 31) are available only from MUSICAM USA, and represent some of the MUSICAM USA enhancements.

8.6.2 Parameter K (10) — Joint Minimum Subband

This parameter ranges from 0 to 31 and represents the minimum subband at which joint stereo is permitted. The ISO specification allows joint stereo to begin at subband 4, 8, 12 or 16. Setting K to 5 would set the minimum to 8. Setting this parameter to 1 would set the minimum subband for joint stereo to 4.

8.6.3 Parameter L (11) Anticorrectional Factor

Unused at present, keep at 1.0 for joint stereo. This parameter attempts to determine if there is a sub-band in which the left and right channels have high levels, but when summed together to form mono, the resulting mono mix has very low levels. This occurs when the left and right signals are anti-correlated. If anti-correlation occurs in a sub-band, joint stereo which includes that sub-band cannot be used. In this case, the joint stereo boundary must be raised to a higher sub-band. This will result in greater quantization noise but without the annoyance of the anti-correlation artifact. A low value of L indicates that if there is a very slight amount of anti-correlation, then move the sub-band boundary for joint stereo to a higher value.

8.6.4 Parameter M (12) Limit Subbands

This parameter can range from 0 to 31 in steps of 1, and limits the allowed number of subbands which receive at least the minimum number of bits. Setting this to 8.3 would insure that sub-bands 0 through 7 would receive the minimum number of bits independent of the psychoacoustic model. It has been found that the psychoacoustic model sometimes determines that no bits are required for a sub-band and using no bits as the model specifies, results in an annoying artifact. This is because the next frame might require bits in the sub-band. This switching effect is very noticeable and annoying. See parameter { for another approach to solving the sub-band switching problem.

8.6.5 Parameter N (13) Threshold of Hearing Table

The audio quality can be adjusted utilizing different threshold of hearing (TOH) tables. There are currently 4 TOH tables, and these can be selected by the user. There now exists the possibility of using different TOH tables for mono, dual mono, joint stereo and stereo at each bit rate and sampling rate combination.

The four TOH tables are:

0 = low bitrate sensitive hearing
1 = low bitrate normal hearing
2 = high bitrate sensitive hearing
3 = high bitrate normal hearing

Parameter 13 is always a four digit number, and each digit represents a different mode. For example:

1023 = use TOH table 1 for mono
use TOH table 0 for dual mono
use TOH table 2 for joint stereo
use TOH table 3 for stereo

8.6.6 Parameter O (14) Stereo/Dual Mono Safety Margin

This parameter ranges from -12 to +12 dB. It represents the safety margin added to the psychoacoustic model results. A positive safety margin means that more bits are used than the psychoacoustic model predicts, while a negative safety margin means to use less bits than the psychoacoustic model predicts. If the psychoacoustic model was exact, then this parameter would be set to 0.

8.6.7 Parameter P (15) Jnt..SKF.ISO

This parameter ranges from 0 to .999999. It is used only if joint stereo is required by the current frame. If joint stereo is not needed for the frame, then this parameter is not used. The parameter p is used in the following equation:

br = demand bit rate * p

If br is greater than the current bit rate (..128, 192, 256, 384), then the ISO method of selecting scale factors is used. The ISO method reduces temporal resolution and requires less bits. If br is less than the current bit rate, then a special method of choosing the scale factors is invoked. This special model generally requires that more bits are used for the scale factors but it provides a better stereo image and temporal resolution. This is generally better at bit rates of 192 and higher. Setting p to zero always forces the ISO scale factor selection while setting p to .9999999 always forces the special joint scale factor selection.

8.6.8 Parameter Q (16) Joint Boundary Adjustment

This parameter ranges from -7 to 7 and represents an adjustment to the subband where joint stereo starts. If the psychoacoustic model chooses 14 for the start of the joint stereo and the Q parameter is set to -3, the joint boundary set to 11 (14 - 3). The joint boundary must be 4, 8, 12 or 16 so the joint boundary is rounded to the closest value, which is 12.

8.6.9 Parameter R (17) Demand Minimum Factor

This value ranges from 0 to 1 and represents the minimum that the demand bit rate is allowed to be. For example, if the demand bit rate mode of bit allocation is used and the demand bit rate is set to a maximum of 256 kb/s and the R parameter is set to .75, the minimum bit rate is 192 kb/s (256 * .75). This parameter should not be necessary if the model was completely accurate. When tuning with the demand bit rate, this parameter should be set to .25 so that the minimum bit rate is a very low value.

8.6.10 Parameter S (18) Stereo/Dual Mono used subbands

This parameter is for stereo and dual mono modes of operation and ranges from 0 to 31, where 0 means use the default maximum (27 or 30) subbands as specified in the ISO specification. If this parameter is set to 15, then only subbands 0 to 14 are allocated bits and subbands 15 and above have no bits allocated.

8.6.11 Parameter T (19) Joint Frames Count

This parameter ranges from 0 to 24 and represents the minimum number of MUSICAM frames (24 millisecond for 48k or 36 ms for 32k) that are coded using joint stereo. Setting this parameter non-zero keeps the model from switching quickly from joint stereo to dual mono. In the ISO model, there are 4 joint stereo boundries. These are at sub-band 4, 8, 12 and 16 (starting at 0). If the psychoacoustic model requires that the boundary for joint stereo be set at 4 for the current frame and the next frame can be coded as a dual mono frame, then the T parameter requires that the boundary be kept at 4 for the next T frames, then the joint boundary is set to 8 for the next T frames and so on. This prevents the model from switching out of joint stereo so quickly. If the current frame is coded as dual mono and the next frame requires joint stereo coding, then the next frame is immediately switched into joint stereo.

The T parameter has no effect for entering joint stereo, it only controls the exit from joint stereo.

8.6.12 Parameter U (20) Peak <.49> RMS

This is a binary parameter. If the value is less than .499, then the psychoacoustic model utilizes the peak value in the subband to determine the number of bits to allocate for that subband. If the parameter is greater than .499, then the RMS value of all the samples in the subband is used to determine how many bits are needed in each subband. Generally, utilizing the RMS value results in a lower demand bit rate and higher audio quality.

8.6.13 Parameter V (21) 3 dB <.499> 6 dB

This parameter is a binary parameter. If the value is less than .499, the 3 dB additional rule is used for tonals. If it is greater than .499, then the 6dB rule for tonals is used. The addition rule specifies how to add masking level for two adjacent tonal maskers. There is some psychoacoustic evidence that the masking of two adjacent tonal maskers is greater (6dB rule) than simply adding the sum of the power of each masking skirt (3dB). In other words, the masking is not the sum of the powers of each of the maskers. See Fig. 8-8.

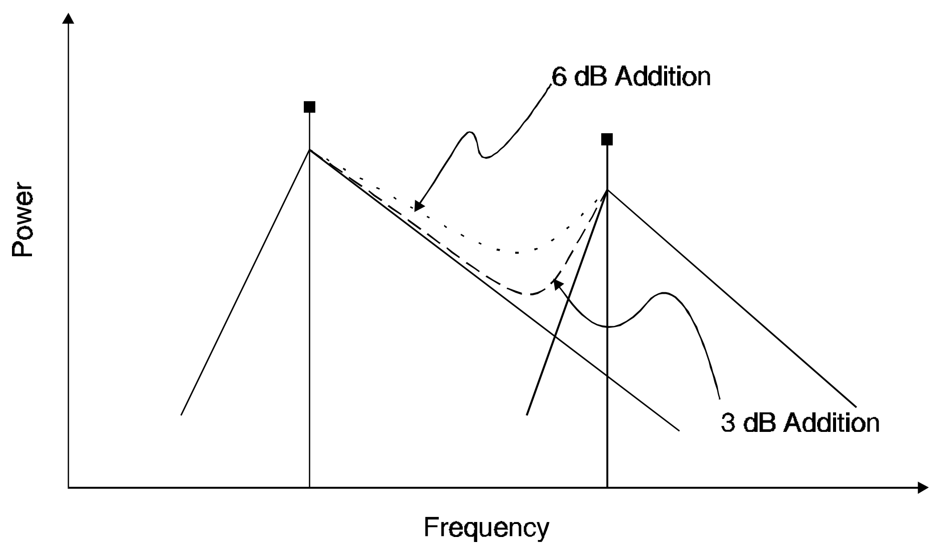


Figure 8-8 Addition of tonal maskers

8.6.14 Parameter W (22) Adjustment Subband 3

This parameter ranges from 0 to 15 dB and represents an adjustment which is made to the psychoacoustic model. It tells the psychoacoustic model to allocate more bits than calculated. A value of 7 would mean that 7dB more bits (remember that 1 bit equals 6 dB) would be allocated to each sample in subband 3. This is used to compensate for inaccuracies in the psychoacoustic model at the frequency of subband 3 (3*750 to 4*750 Hz for 48k sampling).

8.6.15 Parameters X (23) - Z (25) Adjustment Subband 2 - 0

See Parameter W

8.6.16 Parameter { (26) Subband hang Time

The psychoacoustic model may state that a sub-band does not need any bits. The { parameter controls this condition. if the parameter is set to 10, then if the model calculates that no bits are needed for a certain sub-band, 10 consecutive frames must occur with no request for bits in that sub-band before it is turned off. There are 32 counters, one for each sub-band. The { parameter is the same for each sub-band. If a sub-band is turned off, and the next frame needs bits, the sub-band is immediately turned on. This parameter is used to prevent annoying switching on and off of sub-bands. Setting this parameter non-zero results in better sounding audio at higher bit rates but always requires more bits. Thus, at lower bit rates, the increased usage of bits may result in other artifacts.

8.6.17 Parameter | (27) Scale Factor Adjustment

If this parameter is less than .49999, then scale factor adjustments are made. If this parameter is .5000 or greater, then no scale factor adjustments are made (this is the ISO mode). This parameter is used only if joint stereo is used. The scale factor adjustment considers the left and right scale factors as a pair and tries to pick a scale factor pair so that the stereo image is better positioned in the skf plane. The result is that the stereo image is significantly better in the joint stereo mode.

8.6.18 Parameter (28) Mono used subbands

This parameter is for mono modes of operation and ranges from 0 to 31 where 0 means use the default maximum (27 or 30) subbands as specified in the ISO specification. If this parameter is set to 15, then only subbands 0 to 14 are allocated bits and subbands 15 and above have no bits allocated.

8.6.19 Parameter (29) Joint Stereo used subbands

This parameter is for joint stereo modes of operation and ranges from 0 to 31 where 0 means use the default maximum (27 or 30) subbands as specified in the ISO specification. If this parameter is set to 15, then only subbands 0 to 14 are allocated bits and subbands 15 and above have no bits allocated.

8.6.20 Parameter (30) Mono Safety Margin

This parameter ranges from -12 to +12 dB. It represents the safety margin added to the psychoacoustic model results. A positive safety margin means that more bits are used than the psychoacoustic model predicts, while a negative safety margin means to use less bits than the psychoacoustic model predicts. If the psychoacoustic model was exact, then this parameter would be set to 0.

8.6.21 Parameter (31) Joint Stereo Safety Margin

This parameter ranges from -12 to +12 dB. It represents the safety margin added to the psychoacoustic model results. A positive safety margin means that more bits are used than the psychoacoustic model predicts, while a negative safety margin means to use less bits than the psychoacoustic model predicts. If the psychoacoustic model was exact, then this parameter would be set to 0.